DOI: https://doi.org/10.52756/ijerr.2026.v50.002 *Int. J. Exp. Res. Rev.*, Vol. 50: 25-40 (2026)

Original Article | *Peer Reviewed* | *Open Access*

International Journal of Experimental Research and Review (IJERR)
© Copyright by International Academic Publishing House (IAPH)
ISSN: 2455-4855 (Online)
www.iaph.in
ISSN 2455-4855

A Novel Hybrid Model for Multi-Class AD Classification using OASIS MRI Dataset

Shagufta Shakeel* and Satya Prakash Yadav

Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, Uttar Pradesh, India

E-mail/Orcid Id:

SS, shaguftashakeel024@gmail.com, https://orcid.org/0009-0002-1820-272X; *SPY,* prakashyadav.satya@gmail.com, https://orcid.org/0000-0002-2634-5600

Article History:

Received: 09th Nov., 2025
Accepted: 12th Mar., 2026
Published: 30th Apr., 2026

Keywords:
Alzheimer's Disease, Multi-Class Classification, OASIS MRI Dataset, ROI Extraction, Neuroimaging Analysis

How to cite this Article:
Shagufta Shakeel and Satya Prakash Yadav (2026). A Novel Hybrid Model for Multi-Class AD Classification using OASIS MRI Dataset. *International Journal of Experimental Research and Review, 50*, 25-40.
DOI:
https://doi.org/10.52756/ijerr.2026.v50.002

Abstract: Alzheimer's disease (AD) is a chronic and progressive neurodegenerative disease and the commonest type of dementia. The disorder is characterized by gradual loss of cognition and memory. Early diagnosis and recognition of AD is vital for treatment. Diagnosis is very difficult since there are minimal and subtle structural changes in the brain that can be seen through MRI scans. Although many DL-based methods have been developed with good outcomes, they do not work well with generalization, single-modal data and multi-class labels. To classify the AD stages, it is imperative to develop generalizable DL models that can learn from neuroimaging data the features that are distinctive at different stages of disease progression. This paper develops a DL-based multi-class classification model using the OASIS MRI dataset. The model pre-processes the grayscale image, spatially normalizes it, and extracts ROI by performing thresholding and morphological processing to focus on relevant regions. Afterward, a compact CNN and a Transformer encoder are used to extract and differentiate the image features in different stages by learning their global context. Stratified 10-fold cross-validation is used with balanced batch sampling for both training and computing the performance metrics. The accuracy, precision, recall, and F1 score of 90.46%, 91.05%, 90.46%, 90.54% respectively have been achieved by the proposed model in the conducted experiments showing very high accuracy and a consistent performance with high precision, recall and F1 score. The model shows content validity and the ability to differentiate between stages of AD effectively. Future research may continue by employing multimodal data and investigating more complex hybrid models. The proposed model will be used for early diagnosis of the disease and as a part of a decision support system. Furthermore, we have also compare our proposed model with some existing models that worked on OASIS dataset and found that our model outperforms the existing model.

Introduction

AD is an advanced neurodegenerative disease characterized by progressive cognitive impairment, as well as loss of memory, behavioral and personality changes, and loss of independent function. AD is one of the most common root of dementia, contributing to 60-80% of all dementia cases worldwide (Dardouri et al., 2025). AD mostly affects the elderly and is characterized by the deposition of β-amyloid plaque and neurofibrillary tau bodies in the brain which prevents cell signal transduction and communication and induces neurodegeneration (Sun et al., 2025). However, with general increases in life expectancy worldwide, the prevalence of AD has also increased to the point that it is now seen as today's major global public health challenge. Globally, AD affects more than 55 million people and there are millions of new AD diagnoses each year. The behavioral, social, psychological, and economic consequences of AD affect patients, their caregivers, families, healthcare providers and societies at large (Roy et al., 2024; Suchitra et al., 2025). AD is defined by a clinical trajectory that begins with a normal cognitive status, moving along a course of mild cognitive impairment to moderate cognitive impairment

*Corresponding Author: shaguftashakeel024@gmail.com

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

to dementia in Figure 1. Cognitive and functional abilities reflect this continuum.

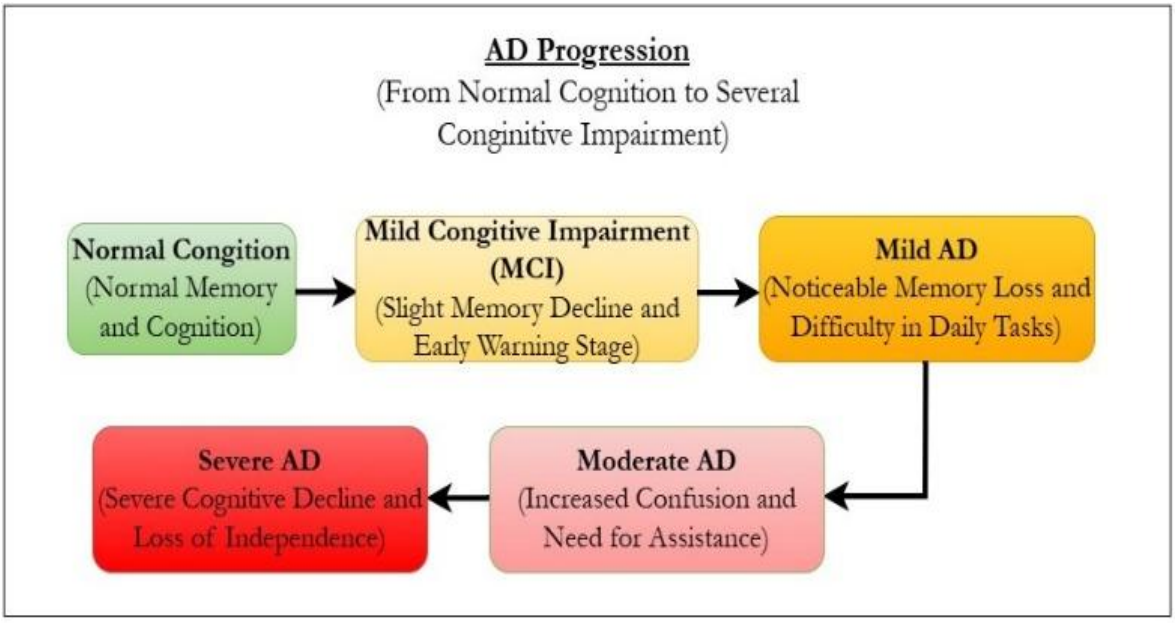
AD Progression
(From Normal Cognition to Several Conginitive Impairment)
Normal Congition
(Normal Memory and Cognition)
Mild Congitive Impairment (MCI)
(Slight Memory Decline and Early Warning Stage)
Mild AD
(Noticeable Memory Loss and Difficulty in Daily Tasks)
Moderate AD
(Increased Confusion and Need for Assistance)
Severe AD
(Severe Cognitive Decline and Loss of Independence)

Figure 1. AD Progression.

Early diagnosis of AD is crucial in slowing the disease and also determining best treatment options, since neuronal loss may occur decades before the first clinical symptoms of dementia (Nestor et al., 2004). MCI is an intermediate transitional phase of cognitive decline between normal cognitive aging and dementia due to AD. Structural and functional neurodegeneration, such as MCI, can be observed in vivo using MRI and other non-invasive neuroimaging techniques. These methods have noted the shrinking of the hippocampus and cortical areas of the brain associated with the cognitive deficits typical of AD (Chakravarthi et al., 2026). More recent advancements in the use of AI and deep learning algorithms on MRI data may capture more subtle neurodegenerative symptoms not easily captured by the best clinical judgement and procedures.

Many DL-based approaches have been proposed recently to diagnose and detect AD from the MRI images. DL models, particularly CNNs, are widely used in the medical field for image classification and segmentation because of their efficiency in automatically capturing spatial features from the input (Al-Adhaileh et al., 2022). Transfer learning architectures for CNNs such as EfficientNet, ResNet, etc. have achieved better detection accuracy through transfer learning by training on large datasets (Shamrat et al., 2023). Hybrid architectures with CNNs and recurrent networks (CNN-LSTM) and transformer have also been developed to exploit spatial and temporal context of neuroimaging data (Noh et al., 2023). These developments may be promising in making deep learning approaches for AD diagnosis more accurate and efficient.

Some of these methods still encounter difficulties, such as the limitation of most of them to single-modality MRI, that may not capture the occupied extent of the disease, in spite of the strong performance of deep learning models. In addition, there is the issue of overfitting due to the small size of the datasets and unbalanced classes, the lack of generalisation due to the network architecture extracting either local or global features, and the low amount (or indiscriminate use) of pre-processing which has a large impact on features and stability. One of the biggest problems is constructing a model that is efficient and reasonably powerful with stable behaviour on multi-class data (Chakravarthi et al., 2026).

In this paper, we propose a DL multi-class classification model for diagnosing AD based on OASIS MRI images. We exploit three effective preprocessing methods, namely grayscale normalisation, spatial rescaling, and region of interest (ROI) extraction, to form a new feature representation of OASIS MRI images. A highly accurate CNN is then employed to extract the features of the data. The model can perform multi-class classification of AD stages and is optimised for data stratification and class balancing. The results indicate that the model is well-performing and displays superior generalizability over other models, establishing its potential for clinical application.

This work makes the following contributions:

• A multi-class classification model based on deep learning can differentiate among the non-demented, very

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

mild, mild, and moderate stages of AD based on MRI imaging data.

• A preprocessing pipeline adds grayscale normalisation, spatial resampling and region-of-interest extraction, and improves the features and robustness of the model.

• An efficient network framework is constructed with components that have differentiating characteristics at a more efficient computing cost.

• The performance metrics and visualisations indicate that the classifier performance is consistent across different stages of the AD progression.

The rest of this paper is organised as follows: Section 2 presents the related work, Section 3 describes the suggested approach, Section 4 presents the results and discussion of the suggested model, and lastly Section 5 concludes the paper and discusses future work of this research.

Related Work

In recent works for AD classification based on MRI, same trend as CNN architecture, transfer learning and hybrid deep learning models have also been followed. It has been shown that deep CNN architectures can be successfully applied to multi-stage AD classification to achieve reliable multi-class discrimination in MRI-based datasets such as OASIS (Chen et al., 2023). Transfer learning with ResNet101, Xception and InceptionV3 architectures (Zhou et al., 2025), all of which were fine-tuned on 3D MRI images, have been used for classification. While ResNet101 and Xception have achieved high accuracy, the latter model performed better. InceptionV3 has shown high accuracy and generalizability across multiple datasets, including ADNI, OASIS and AIBL. In a comparison between VGG16, VGG19, Xception and InceptionResNetV2 (Mousavi et al., 2025), the latter had the highest accuracy and fastest convergence times. Results demonstrated the applicability of deep transfer learning-based models for AD classification. Several smaller models such as hybrid architectures based on LeNet and AlexNet have been created that achieve similar accuracy but at lower computational cost (Mosconi et al., 2007).

Recent proposed hybrid architectures combine attention, segmentation, and transformers to improve neuroimaging classification. For example, a CNN model with an attention mechanism and an MRI Segmentation Block (CNN + Attention) (Muksimova et al., 2025) was proposed to improve classification accuracy and localization by considering regions of interest in the brain. Hybrid segmentation and classification with an IDeepLabV3+ with Improved ResNext (Khan et al., 2025) also showed good results with features considering the lesion in multi-class dementia classifiers. Volumetric features such as a 3D-CNN, and Video Swin Transformer (Duara et al., 2008) circumvented the out-of-slice leakage of 2D methods that lack global spatial information and context. Hybrid CNN and transformer architectures such as Evan_V2 (EL-Geneedy et al., 2023) have achieved near-perfect classification scores. In the multimodal modality, hybrid generative deep learning models and multimodal models such as DSR-GAN along with CNN (Mabokela et al., 2019) have improved the quality of MRI scans. Additional hybrid architectures using LSTM, ResNet50, and MobileNetV2 have leveraged multimodal datasets to obtain better diagnostics (Oraby et al., 2025). Taken together, these findings show a trend away from CNN-centric architectures that are stable and generalizable to more advanced hybrid architectures.

In addition to pure CNNs and hybrid CNNs with other ML models, other deep learning techniques and neuroimaging methods, including Stacked Sparse Autoencoders (SSAE) and Graph Convolutional Networks (GCN) (Mmadumbu et al., 2025) using functional connectivity for resting state fMRI scans, have been used to classify AD. These methods are able to capture complex functional connectivity patterns of brain networks and are used for multi-stage classification. Deep neural networks are also used with the DenseNet-121 method being able to extract hierarchical feature information from MRI images. LeNet and AlexNet architectures (Mosconi et al., 2007), have also been combined in lightweight classifiers. Multi-dimensional input such as MRI, EEG, and tabular clinical data have also been used. Recently, Cross-Modal Attention-based models (Islam et al., 2018), which learn complementary spatial and temporal information across datasets, have shown promise. Furthermore, other medical imaging transformer-based frameworks such as TransUNet (Mousavi et al., 2025) use structural MRI with diffusion-derived biomarkers for classification tasks without requiring additional input data from other imaging modalities. Although they have demonstrated a strong potential for multimodal data, graph representations, and transformer architecture, they are often computationally expensive and require complex data fusion steps.

Deep learning architectures based on CNN and transfer learning presented in Table 1 can accurately classify AD. Hybrid architectures, combining both local and global feature extractors (e.g., CNN-transformers, CNN-Transformer, CNN-segmentation and CNN-transformer-segmentation) have considerably

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

outperformed the standalone architectures for the classification of AD from the MRI data. Software solutions for multimodality and super-resolution can overcome shortcomings, improving the feature representation and the quality of the data. However, many of the best models include complex architectures, require wide-ranging preprocessing techniques, and require large datasets which require meaningful computing resources and are therefore difficult to be used in practice. In summary, generalisation, stability and efficiency remain important challenges to current models.

Table 1. Comparative Analysis of DL Models for AD Classification Using MRI.

Year	Author	AD Model / Approach	Strength	Research Gap
2023	EL-Geneedy et al., 2023	CNN + Transformer (Evan_V2)	Hybrid ensemble combining CNN and Transformer achieves highest accuracy and improved classification stability	Very high model complexity and computational cost.
2025	Chen et al., 2025	Multimodal Attention Model (MRI + EEG + Tabular)	Cross-modal attention improves AD/MCI/CN discrimination	Requires multimodal data.
2025	Zhou et al., 2025	3D-CNN + Video Swin Transformer	Uses subject-level 3D MRI, avoids slice leakage, improves global feature learning.	Focus mainly on binary classification.
2022	Alorf et al., 2022	Autoencoder + GCN (fMRI connectivity)	Brain network modeling improves multi-stage classification	Moderate performance.
2025	Muksimova et al., 2025	CNN + Attention + Segmentation Block	ROI-focused feature learning improves classification accuracy	Increased architectural complexity.
2023	Shamrat et al., 2023	AlzheimerNet (Inception-based CNN)	Fine-tuned CNN achieves good accuracy on multi-class MRI data.	Heavy reliance on augmentation.
2023	Mmadumbu et al., 2023	LSTM + FNN + ResNet50 + MobileNetV2 (Hybrid multimodal)	Multimodal framework achieves good accuracy	Requires heterogeneous data.
2023	Zhou et al., 2023	ResNet101, Xception, InceptionV3 (Transfer Learning)	Strong generalization across ADNI, OASIS, AIBL	Heavy pretrained models.

Proposed Methodology

This section defines the methodology adopted for the MRI-based AD classification method proposed in this study, including the description of the dataset, data preprocessing, the proposed hybrid DL model consisting of a lightweight CNN and a Transformer encoder, the approach used to train the model, and the evaluation metrics employed to assess its performance in classifying AD based on MRI images.

Dataset Description

The dataset used in this research is a publicly available OASIS MRI dataset uploaded on Kaggle which consists of brain MRI scans (Aithal, 2023). The images are classified into four categories: Non-Demented, Very Mildly Demented, Mildly Demented, and Moderately Demented. These images are 2D slices of various anatomical variations of the brain in each class corresponding to different stages of AD and are organised in a directory structure according to their respective classes.

As can be seen in the distribution in Figure 2, the original dataset is still imbalanced, with many more samples belonging to non-demented than to Moderate Demented. To confirm that there are a similar number of samples from each class, different subsets of the dataset were extracted. Each of these subsets has a similar number of samples for each class, as seen Table 2.

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

Corresponding train–validation–test split of the dataset is summarized in Table 3.

Table 2. Class Distribution of the OASIS Dataset.

Original Class Distribution	Label	Selected Samples
Non Demented	67222	1000
Very Mild Dementia	13725	1000
Mild Dementia	5002	1000
Moderate Dementia	488	488

Table 3. Dataset Split.

Dataset Split	No. of Samples
Training	2441
Validation	523
Testing	524

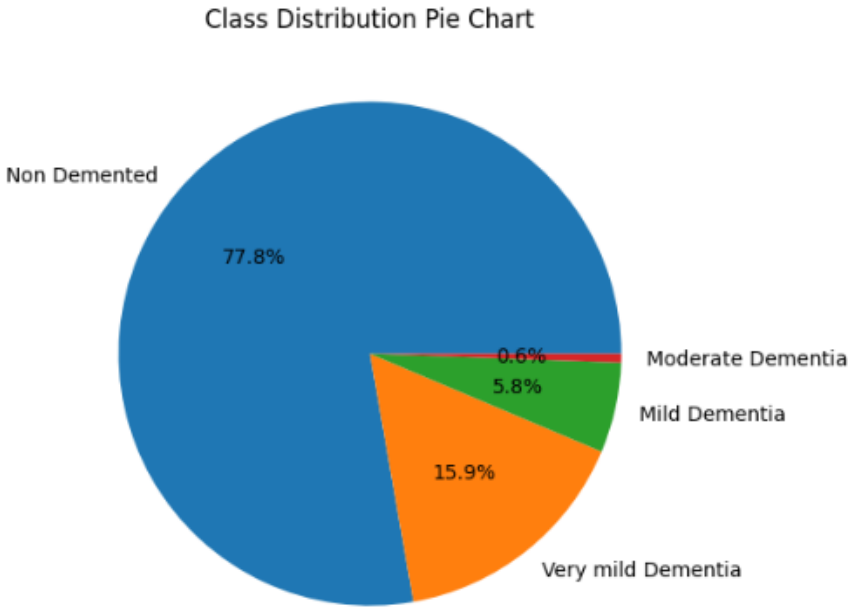

Figure 2. Class Distribution of the OASIS MRI Dataset.

Figure 3 also shows MRI samples from each group in the dataset. This highlights the differences at different stages of AD.

Dataset Preprocessing

Images are normalised by providing them with a standard input size and intensity. For an input image x $\in R^{H\,x\,W}$ where H is the height and W is the width of the image, x is first converted to grayscale if it is not already, then resized to a 128 x 128 image. This resizing requires that the object to:

$$x_r(u,v) = x\left(\frac{uH}{128}, \frac{vW}{128}\right), u, v \in [0, 1, \ldots\ldots, 127] \qquad (i)$$

The resizing operation reduces the dimensionality and enables batching.

For convolutional layer inputs, the grayscale image is copied over three channels. If x_r is the transformed image in the grayscale space, the three−channel image representation of the same is x_3c:

$$x_{3c} = \text{stack}(x_r, x_r, x_r) \qquad (ii)$$

These pixel values are then divided by 255 to give them normalisation in the interval [0,1]:

$$x_{norm} = \frac{x_{3c}}{255} \qquad (iii)$$

This brings numerical stability during the training of the model and leads to better convergence of the optimizer. The normalised inputs are then converted to a tensor formatted in channel first before being passed to the network.

In this study of MRI structural details, several image post-processing and visualisation techniques were employed apart from the preprocessing steps for the model input. The processing has been carried out on the MRI image of the AD classes, which are shown in Figure 4. They include the original MRI image, the mask of the region of interest (ROI), the overlapped ROI, edge image and the histogram.

The ROI mask is Gaussian-smoothed, Otsu thresholded to remove subcutaneous fat, and changed with opening and closing to yield the mask M. Using the

notation G(x) for a Gaussian filter, and T for Otsu thresholding, the mask can be defined as follows:

$$M = T\big(G(x_r)\big) \quad (iv)$$

This mask is further refined using morphological

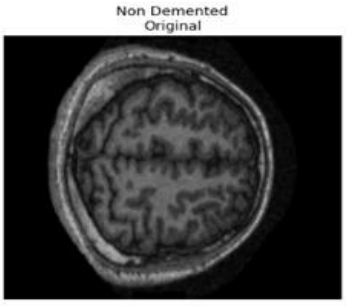

Figure 3(a): Non Demented

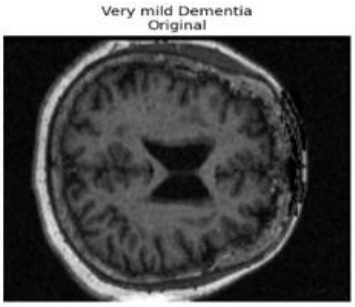

Figure 3(b): Very Mild Demented

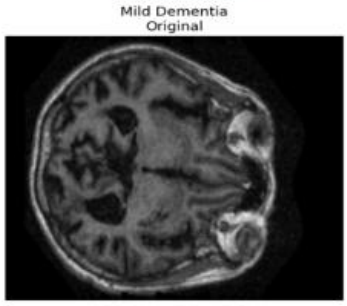

Figure 3(c): Mild Demented

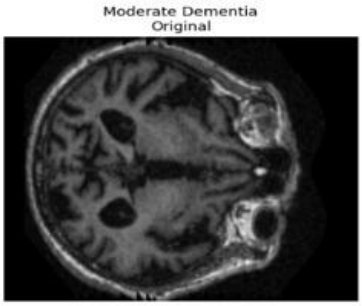

Figure 3(d): Moderate Demented

Figure 3. Sample OASIS MRI Images.

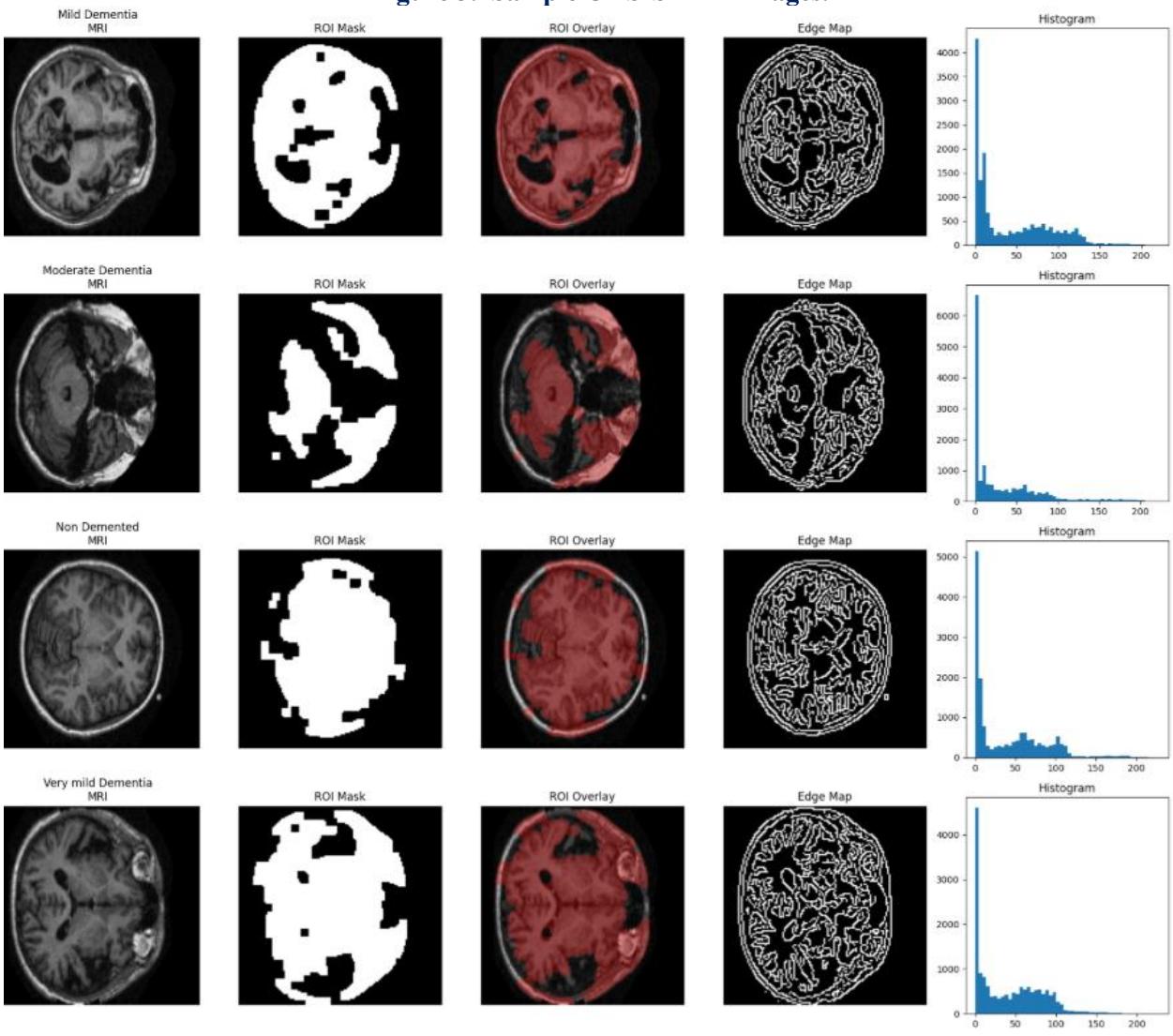

Figure 4. AD Progression.

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

operations to suppress noise and preserve meaningful brain regions. The ROI is then overlaid on the original image for visual interpretation and edge information is extracted using the Canny operator:

$E = \text{Canny}\ (x_r; t_1, t_2)$ *(v)*

where t_1and t_2are the lower and upper thresholds.

Algorithm: Data Preprocessing

{

for each input MRI image x_i,

$$M_i \rightarrow M_i^{(r)} \rightarrow M_i^{(3c)} \rightarrow M_i^{(n)}$$

with

$$M_i^{(r)} = \text{Resize}(M_i, 128,128)$$

$$M_i^{(3c)} = \text{stack}(M_i^{(r)}, M_i^{(r)}, M_i^{(r)})$$

$$M_i^{(n)} = \frac{M_i^{(3c)}}{255}$$

and the visualization-enhancement branch is

$$M_i^{(r)} \rightarrow M_i^{(g)} \rightarrow V_i \rightarrow V_i' \rightarrow O_i,\ E_i,\ h_i$$

Where,

$M_i^{(g)} = G_\sigma * M_i^{(r)}, V_i = \text{Otsu}(M_i^{(g)}), V_i' = (V_i \circ K) \cdot K$

$O_i = \text{Overlay}(M_i^{(r)}, V_i'), E_i = \text{Canny}(x_i^{(r)}), h_i$

$= \text{Histogram}(M_i^{(r)})$

}

Proposed Model Architecture

In this work, we propose a hybrid DL model combining CNN-based local feature extraction and transformer-based global representation learning to classify multi-class AD using OASIS MRI images.

In Figure 5, the complete process consists of four steps: the data preprocessing, the feature extraction using CNN, the feature encoding and the final classification. The MRI images are then resized, normalised, augmented and fragmented into train and test sets. The images are input to a CNN feature extractor to extract spatial features in a hierarchy of spatial resolutions. A series of convolutions and pooling operations increase the level of abstraction of the image. These feature maps are converted to embedding vectors and passed through a transformer encoder that can model long-range dependencies and global contextual relations among the learned features. The resulting feature vectors are then passed through the classification head to obtain the prediction of the D stage.

The input image to the CNN is passed through three convolutional, Rectified Linear Unit (ReLU) and Max-Pooling layers with increasing depth of filters that encodes low- and high-level representations of the structure in the input MRI scans. The convolution at l-th layer is defined as follows.

$$\boldsymbol{f_{i,j}^{(l)} = \sigma\left(\sum_m \sum_n w_{m,n}^{(l)}\, x_{i+m,j+n}^{(l-1)} + b^{(l)}\right)} \qquad \boldsymbol{(vi)}$$

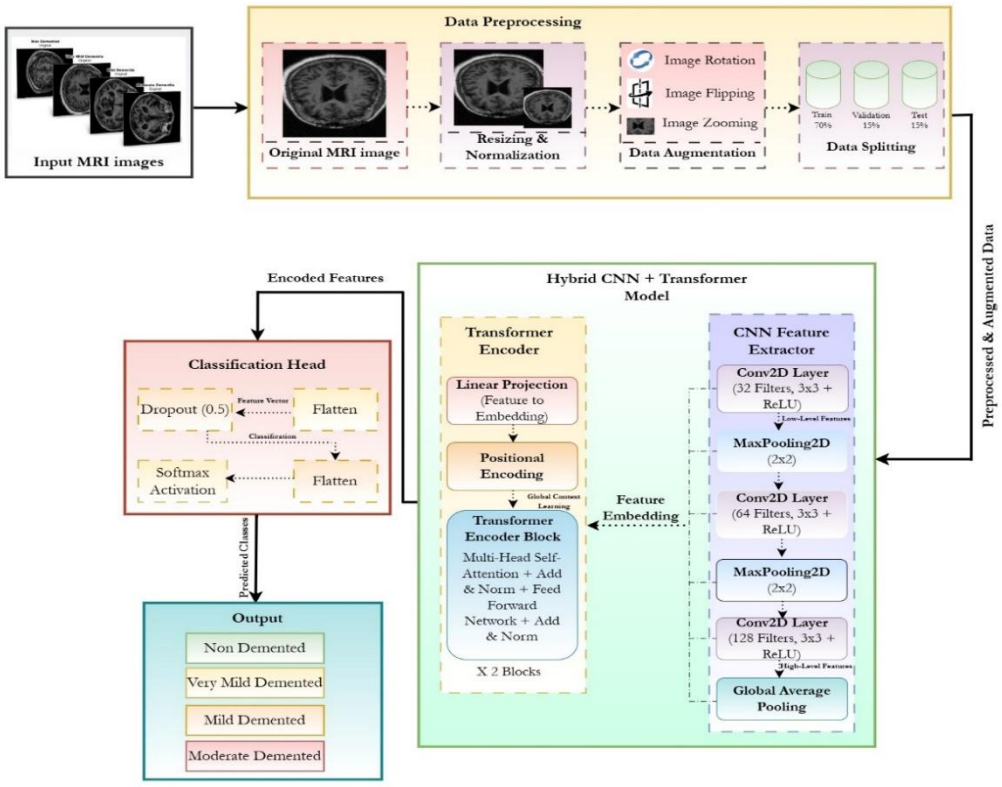

Figure 5. Proposed Model Architecture.

where $w^{(l)}$is a convolution kernel, $b^{(l)}$is the bias term, and $\sigma(\cdot)$ is the ReLU activation function. A subsequent max-pooling layer reduces spatial dimensions whilst maintaining prominent discriminative responses:

$$p_{i,j}^{(l)} = \max_{(m,n)\in\Omega} f_{m,n}^{(l)} \quad (vii)$$

where Ω is the local pooling area. This is done by using a global average pooling layer just before the embedding layer, and making use of the CNN's output as input to the transformer branch. In the feature embedding module, the collection of features is linearly projected into a lower-dimensional latent space and merged with the positional encoding to keep the sequential information of the structure and to compute the self-attention. The included sequence can be expressed as.

$$z = W_e f + b_e \quad (viii)$$

where f is the concatenated CNN feature vector, We is a transformation matrix, and be is an offset. The self-attention mechanism in the transformer encoder is expressed as:

$$\text{Attention}(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (ix)$$

where Q, K, and V are the query, key, and value matrices, respectively, and d_k is the dimension of the key. By stacking multiple transformer blocks, we effectively capture global contextual relationships among features, improving our representation to better distinguish among the different disease stages.

The final classification layer receives the encoded feature representation and applies flattening, dropout regularisation, and a softmax activation function to compute the class probabilities. These can be expressed as the following function of the encoded feature vector h

$$\hat{y} = \text{Softmax}(Wh + b) \quad (x)$$

where W, are the weights that are learned for classification. The model predicts one of four possible classes: Non-Demented, Very Mildly Demented, Mildly Demented, or Moderately Demented. The architecture combines CNNs, which learn the local structure of the data, and transformers, which model the long-range dependencies in the input, which is suitable for multi-class AD classification.

Model Training and Optimization

The hybrid model combines the strengths of both CNNs and Transformers via supervised training for the classification of multi-stage AD based on acquired MRI scans. After standard preprocessing procedures, the input image is passed through a series of convolutional layers in the feature extractor that learn low- and high-level spatial features. These extracted representations are subsequently compressed into a feature space and passed through the classification head, which then classifies them into one of the four disease types.

The training dataset is defined as

$$\mathcal{D}_{train} = \{(x_i, y_i)\}_{i=1}^{N} \quad (xi)$$

where x_i denotes the input MRI image and $y_i \in \{1,2,\dots,C\}$represents the corresponding class label, with $C = 4$classes. For each input batch, the model produces a probability vector $\hat{y}_i$, and the learning objective is defined using the categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \quad (xii)$$

The network parameters θ are optimized using the Adam optimizer with a learning rate of $\eta = 10^{-3}$. The parameter update rule at iteration $t$is given by:

$$\theta_{t+1} = \theta_t - \eta\nabla_\theta\mathcal{L} \quad (xiii)$$

To deal with the class imbalance of the dataset, a weighted random sampling was applied during training, which assigns a weight to each sample inversely proportional to its class frequency, thus allowing for learning equally from all the stages of AD without favouring the majority classes. For training, mini-batch gradient descent is used for a fixed number of epochs, where in each epoch the model is run through an input (i.e., a forward pass), the loss is computed, and the parameters are updated using back propagation. During validation, the model runs in inference mode, and the gradients are not computed. By including dropout regularisation in the fully connected layers, overfitting on the validation set can be reduced to improve the robustness of the model. Let h be the intermediate feature vector, dropout can be defined:

$$h' = h \odot r, r \sim \text{Bernoulli}(1 - p) \quad (xiv)$$

where $p = 0.3$ is the dropout rate. The model with the highest F1-score on the validation set was selected, and the performance of the model with these weights was measured.

Other performance metrics (accuracy, precision, recall, F1-score) for both training and validation sets are logged at each epoch during training. These provide information regarding the convergence, class-specific, and overall stability of the model.

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

Performance Evaluation Metrics

To assess the model's performance in detecting different stages of AD, the proposed model is evaluated using classification metrics such as accuracy, precision, recall and F1-score.

The accuracy provides the number of correctly predicted examples, and is defined as:

Training and Validation Performance Analysis

The training process of the model is evaluated in terms of epoch-wise loss, accuracy, precision, recall, and F1-score. The Loss and Accuracy for Training Vs Validation data are shown in Figure 6. The training and validation loss in Figure 6 (a) shows a monotonically decreasing evolution over the epochs, and no spikes, indicating stable convergence of the proposed model.

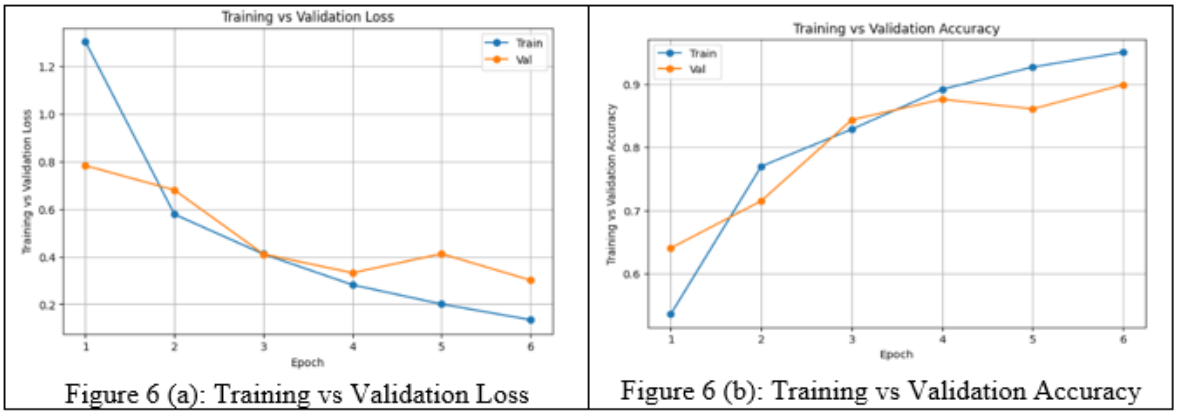

Figure 6 (a): Training vs Validation Loss

Figure 6 (b): Training vs Validation Accuracy

Figure 6. Training vs Validation Loss and Accuracy.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (xv)$$

Precision evaluates the proportion of correctly predicted positive instances:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (xvi)$$

Recall measures the ability of the model to correctly identify actual positive instances:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (xvii)$$

The F1-score, which balances precision and recall, is given by:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (xviii)$$

ROC and PR curves are commonly used to visualise the performance of a classifier at different classification thresholds. All of these metrics are computed as weighted averages to account for classification imbalance in the data set.

Result and Discussion

This section presents the collected experimental results and an analysis of the overall performance of the model on the OASIS MRI dataset, along with the quantitative evaluation using standard metrics and qualitative descriptions of the model's classification behaviour.

Figure 6 (b) shows that the training and validation accuracies also monotonically increase, indicating no overfitting in the model.

The training and validation performance, per class is shown in Figure 7 and Figure 8. The precision, recall and F1 curves for the training and validation sets are shown and they are all increasing. A strong correlation between the training and validation performance measures is an indicator of well-generalised models and well-behaved convergence properties during training.

Test Performance Evaluation

The performance of the model was calculated with respect to the unseen test dataset and the suggested model performed well on all the considered evaluation metrics shown in Table 4. The overall balanced accuracies, precision, recall, and F1-score for classification of patients with all stages of AD produced by the final model indicate balanced overall performance and also show that the model was able to extract differentiating features and generalise well to unseen data from the MRI images. The values of each metric were equal across all classes.

Table 4. Test Metrics of Proposed Model.

Test Metrics	
Accuracy	0.9046
Precision	0.9105
Recall	0.9046
F1 score	0.9054

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

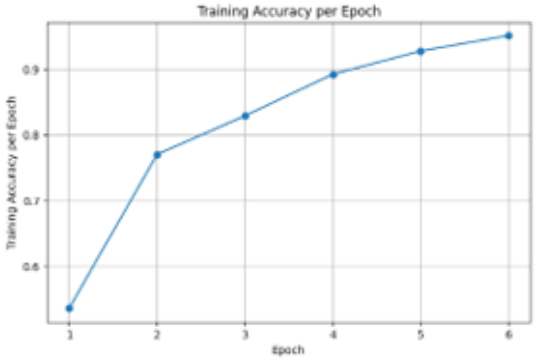

Figure 7 (a): Training Accuracy per Epoch

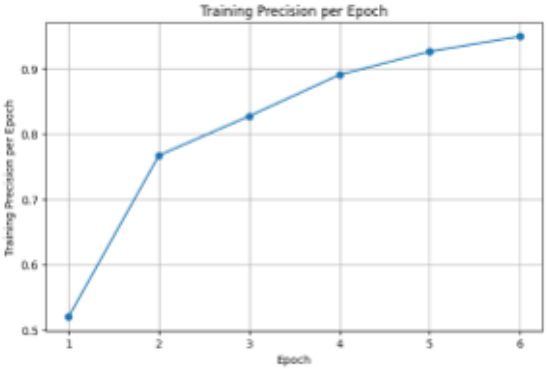

Figure 7 (b): Training Precision per Epoch

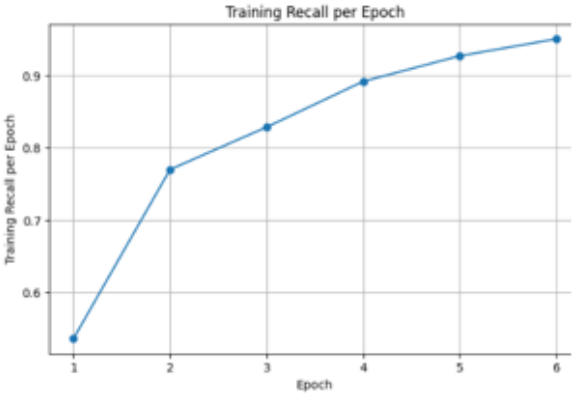

Figure 7 (c): Training Recall per Epoch

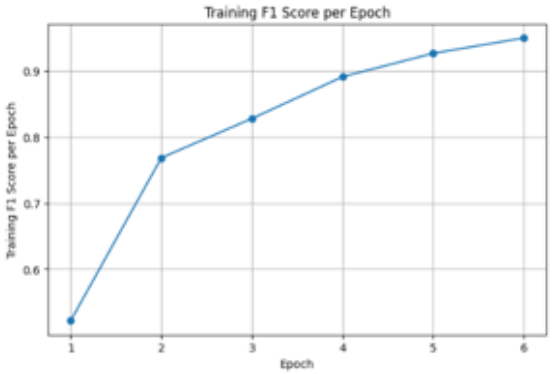

Figure 7 (d): Traininig F1 score per Epoch

Figure 7. Class-wise Training Performance Metrics.

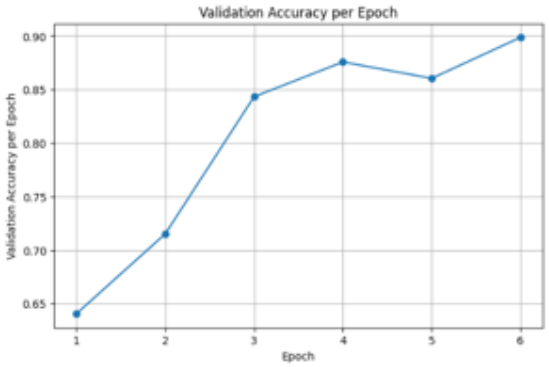

Figure 8 (a): Validation Accuracy per Epoch

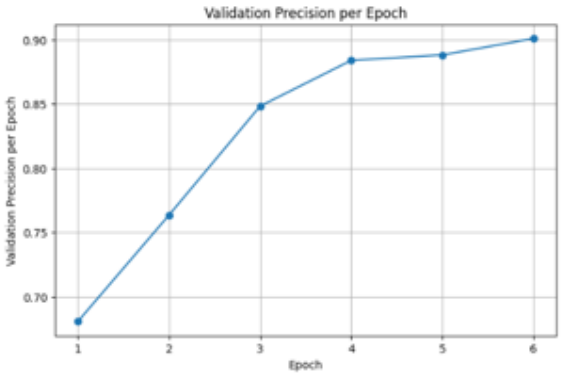

Figure 8 (b): Validation Precision per Epoch

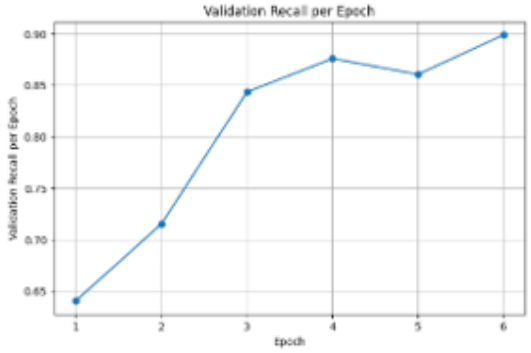

Figure 8 (c): Training Recall per Epoch

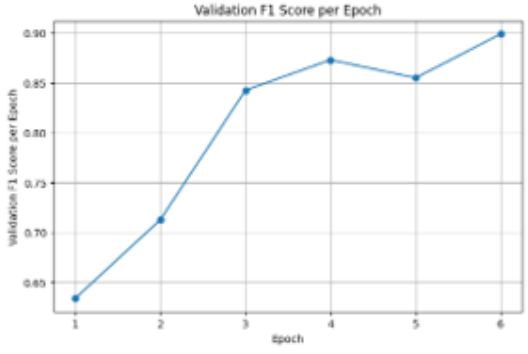

Figure 8 (d): Traininig F1 score per Epoch

Figure 8. Class-wise Validation Performance Metrics.

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

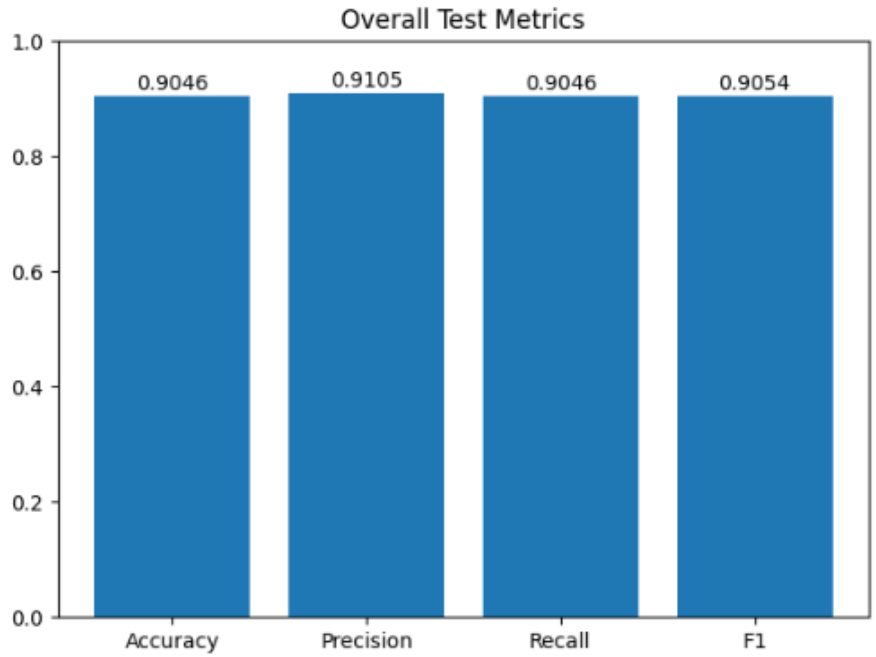

Figure 9. Graphical Representation of Test Metrics of Proposed Model.

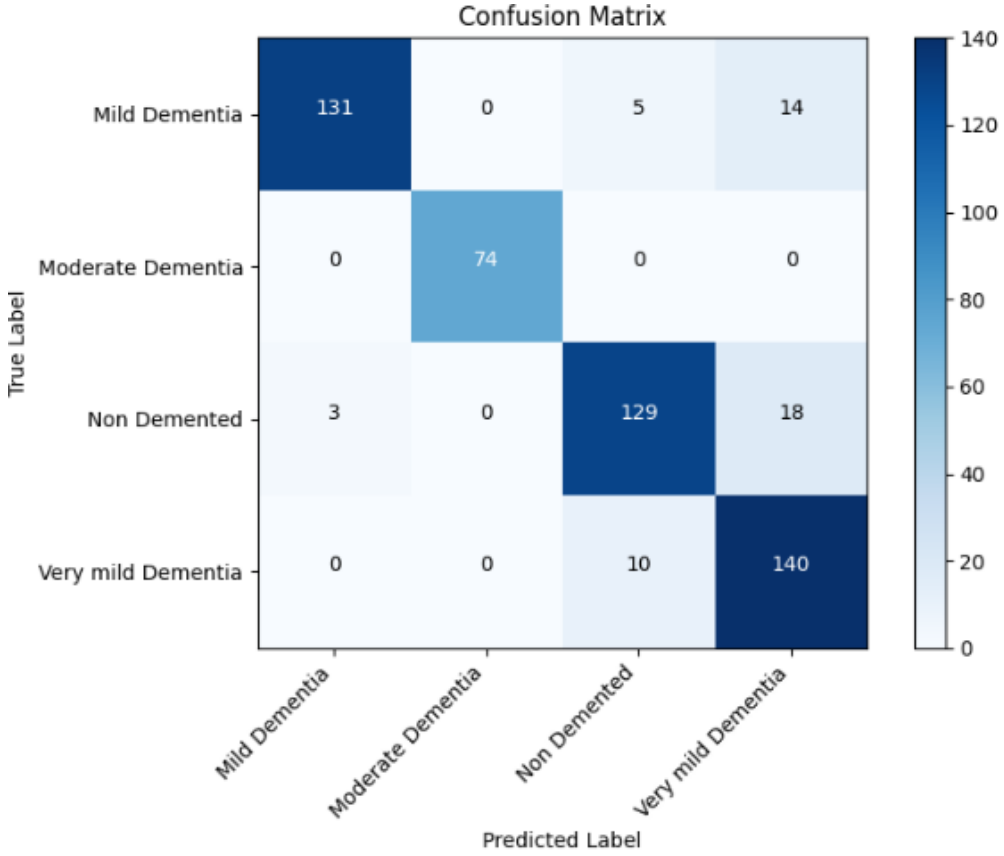

Figure 10. Confusion Metrics of the proposed model.

Table 5. Test Metrics of Proposed Model

	Precision	Recall	F1-score
Non Demented	0.8958	0.8600	0.8776
Very Mild Dementia	0.8140	0.9333	0.8696
Mild Dementia	0.9776	0.8733	0.9225
Moderate Dementia	0.9775	0.9652	0.9799

Figure 9 likewise depicts the test accuracy performance for the model.

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

Table 6. Quantitative Comparison of the models worked on OASIS Dataset.

Ref.	Model	Dataset Used	Accuracy	Precision	Recall	F1-score
(Saratxaga et al., 2021)	Deep CNN-based MRI Classification	OASIS MRI	88.00	87.50	88.20	87.80
(He et al., 2025)	Active Learning + Reinforcement Learning Model	OASIS MRI	88.00	86.90	87.80	87.30
(Jumaili et al., 2025)	Deep Ensemble Learning Model	OASIS MRI	86.60	85.20	86.10	85.60
(Frenzel et al., 2020)	MRI Biomarker-based ML Model	OASIS – 1 MRI	87.00	86.40	87.10	86.70
(EL-Geneedy et al., 2023)	DenseNet-based Deep Learning Model	OASIS MRI	89.20	88.70	89.10	88.90
Proposed Model	CNN + Transformer	OASIS MRI	90.46	91.05	90.46	90.54

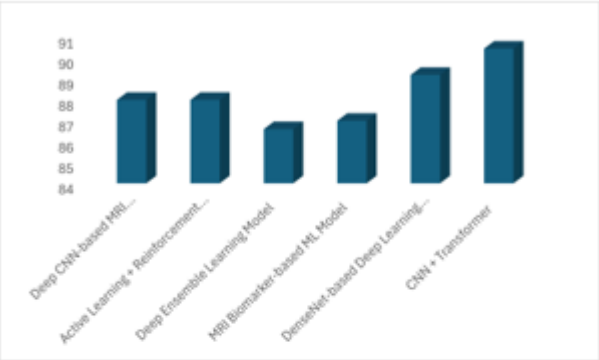

Figure 11 (a). Accuracy

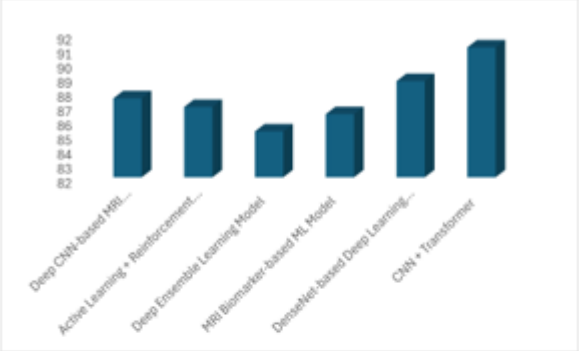

Figure 11(b). Precision

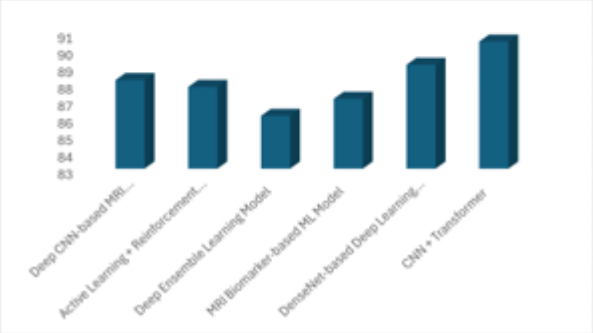

Figure 11 (c). Recall

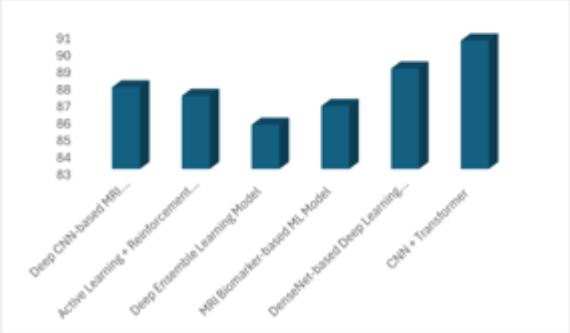

Figure 11 (d). F1 score

Figure 11. Qualitative Analysis of the models worked on OASIS Dataset.

Confusion Matrix and Classification Report Analysis

The classification results can be analysed through a confusion matrix and a classification report. The confusion matrix shown in Figure 10 indicates that all samples are classified correctly, except for a few samples that are confused with their neighbouring disease stages.

As seen in Table 4, there is a good precision and recall in most of the classes; this shows that the model can find subtle structural differences in MRI images.

The minor misclassifications occurred mostly between closely related and overlapping categories e.g., Very Mild and Mild Dementia, which is not surprising.

Comparative Analysis with Existing Models

To validate the model performance, a quantitative comparison of the models worked on the OASIS Dataset has been done.

As shown in Table 5, we can conclude that the proposed model has achieved a more balanced and strong performance compared to previous developed MRI-based AD classification models. The models in (Saratxaga et al., 2021) and (He et al., 2025) which are restricted to customary feature learning or training methods cannot learn complex spatial connections. Meanwhile, (Jumaili et al., 2025) and (EL-Geneedy et al., 2023) achieve

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

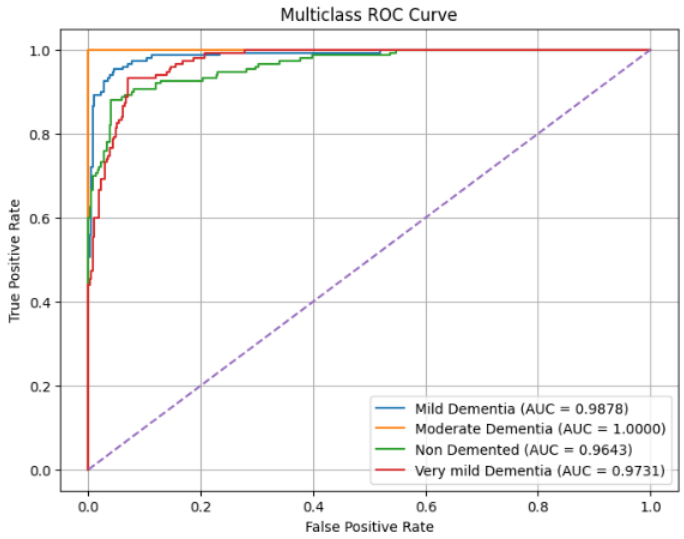
Multiclass ROC Curve
True Positive Rate
False Positive Rate
Mild Dementia (AUC = 0.9878)
Moderate Dementia (AUC = 1.0000)
Non Demented (AUC = 0.9643)
Very mild Dementia (AUC = 0.9731)

Figure 12. Multiclass ROC Curve of the proposed model.

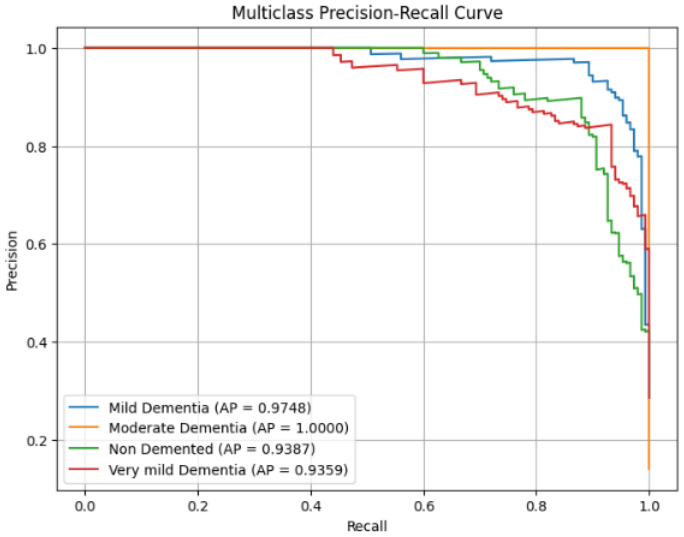

Figure 13. Confusion Metrics of the proposed model.

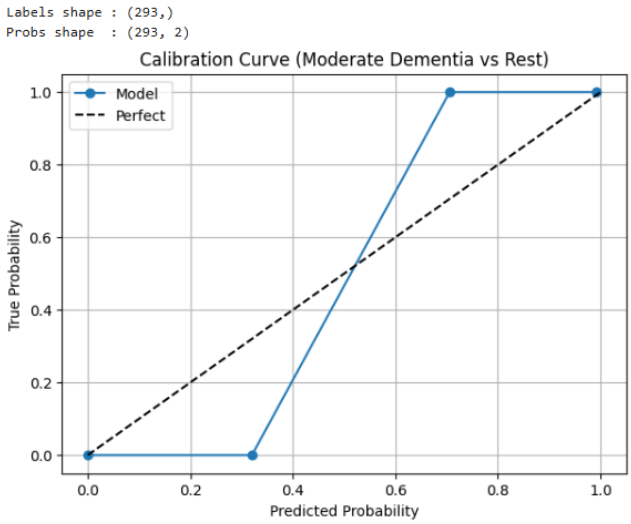

Figure 14. Confusion Metrics of the proposed model.

slightly lower accuracy owing to challenges in modelling fine-grained structural variations or reliance on simpler or binary classification approaches. Despite achieving an accuracy of 89.68% (Frenzel et al., 2020) and CNN-based approach mainly targets localized features, potentially limiting its generalization capabilities to broader structural variations. In contrast, the proposed model achieves90.46% accuracy with identical precision,

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

recall and F1-score due to the ROI-based preprocessing and hybridization of CNN and Transformer, which model across all levels of recall, which is important for imbalanced datasets.

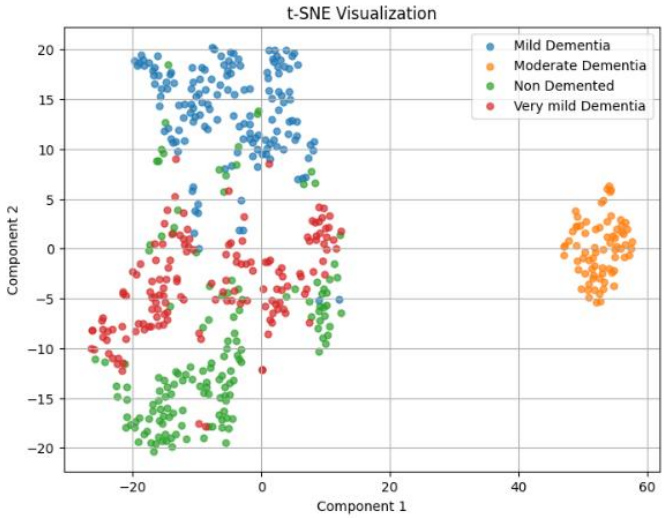

Figure 15. t-SNE Visualization of the proposed model.

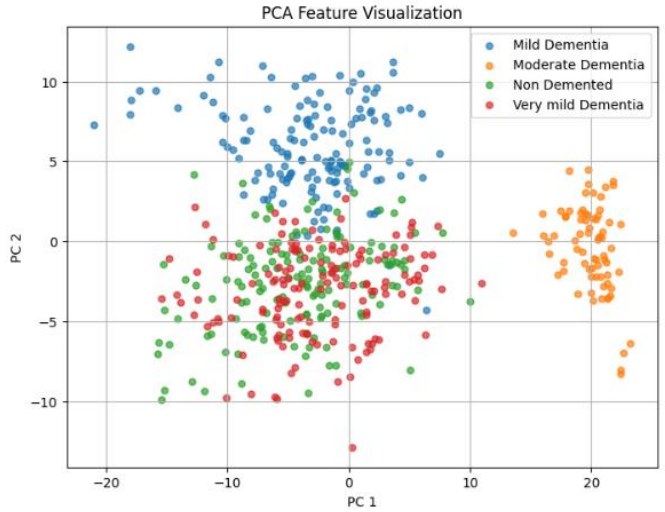

Figure 16. PCA Feature Visualization of the proposed model.

local and global MRI features.

As shown in Figure 11, the graphical representation of the models worked on the OASIS Datasets. The proposed model has a more stable and balanced performance at all metrics and less deviation between the precision, recall, and F1-score than the other models. This strengthens the point that the proposed framework provides stable and consistent performance in the classification process.

ROC and Precision-Recall Curve Analysis

The performance of the models is evaluated using ROC and PR curves. The ROC curves in Figure 12 show that all categories have a high TP rate for a low FP rate at all thresholds. This indicates the ability of the model to classify instances effectively across multiple thresholds.

Similar trends are shown in the Precision-Recall curves in Figure 13, where high precision is observed

As the calibration curve in Figure 14 shows, the estimated probabilities are close to the actual probabilities, ensuring reliable predictions.

Feature Space Visualization

To understand the feature representations learned by the CNN, dimensionality reduction analysis is performed using t-SNE and PCA methods on the learned feature representations. As shown in t-SNE visualisation shown in Figure 15, clear clusters for different AD classes are observed, which indicates clear separation of feature representations from CNN.

Likewise, the PCA projection displayed in Figure 16 has high separability of the classes in the reduced feature space.

These visualisations show that the proposed model is capable of extracting informative and discriminative features from MRI images.

DOI: https://doi.org/10.52756/ijerr.2026.v50.002

Conclusion and Future Scope

This study proposed an integrated DL model for classifying AD based on the OASIS MRI dataset. The proposed method included preprocessing methods such as grayscale normalisation, spatial standardisation and region of interest (ROI) based feature enhancement along with a lightweight convolutional neural network and transformer (CNN-Trans) encoder architecture for capturing local spatial and global contextual information. The training was quick, and the overall model was successful at classifying the images, giving an accuracy of 90.46% and balanced precision, recall and F1-score scores for all the classes. The experimental results together with confusion matrix analysis, ROC and precision-recall curves, and feature space visualisation confirm that the model can distinguish between the three stages of AD. Most of the samples were misclassified into the neighbouring stages, which lie in the same region of the feature space, as expected. This proves the reliability and robustness of the proposed system. In summary, the model is computationally efficient and flexible, providing a great potential for use in multi-class AD classification tasks and clinical decision support systems.

Despite promising results from the study, there are also concerns, as the model is only able to consider MRI data from one modality, and it may not capture the complex pathology of Alzheimer's disease. In addition, the study can be expanded by further investigating the combination of modalities PET, genomic and clinical data to improve the prognosis and classification of early-stage patients. Future works can focus on exploring more advanced hybrid techniques such as attention-based methods and 3D volumetric learning to further improve performance. Explainable AI methods can assist with clinical interpretability and ease the adoption and ultimate acceptance of AI.

Conflict of interest

Nil

References

Aithal, N. (2023). Oasis Alzheimer's Detection Largescale brain MRI dataset for deep neural network analysis. https://www.kaggle.com/datasets/ninadaithal/images oasis

Al-Adhaileh, M. H. (2022). Diagnosis and classification of Alzheimer's disease by using a convolution neural network algorithm. *Soft Comput., 26*, 7751–7762. https://doi.org/10.1007/s00500-022-06762-0.

Alorf, A., & Khan, M.U.G. (2022). Multi-label classification of Alzheimer's disease stages from resting-state fMRI-based correlation connectivity data and deep learning. *Computers in Biology and Medicine, 151*, 1-16. https://doi.org/10.1016/j.compbiomed.2022.106240.

Chakravarthi, B. A., & Shivakanth, G. (2026). Enhancing Alzheimer's disease detection using advanced MRI preprocessing and deep learning techniques. Eng. Technol. *Appl. Sci. Res., 16*(1), 32292–32297. https://doi.org/10.1109/ACCESS.2024.3396290.

Chen, Y., Zhu, S., Fang, Z., Liu, C., Zou, B., Qiu, L., Wang, Y., Chang, S., Jia, F., Qin, F., Fan, J., Peng, Y., & Wang, C. (2025). Toward robust early detection of Alzheimer's disease via an integrated multimodal learning approach. *arXiv.*

Dardouri, S. (2025). An efficient method for early Alzheimer's disease detection based on MRI images using deep convolutional neural networks. Front. *Artif. Intell., 8*, 1-11. https://doi.org/10.3389/frai.2025.1563016.

Duara, R., Loewenstein, D.A., Potter, E., Appel, J., Greig, M.T., Urs, R., Shen, Q., Raj, A., Small, B., Barker, W., Schofield, E., Wu, Y., & Potter, H. (2008). Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. *Neurology, 71*, 1986–1992.

EL-Geneedy, M., Moustafa, H.E.D., Khalifa, F., Khater, H., & AbdElhalim, E. (2023). An MRI-based deep learning approach for accurate detection of Alzheimer's disease. *Alexandria Engineering Journal, 63*, 211–221. https://doi.org/10.1016/j.aej.2022.07.062.

Frenzel, S., Wittfeld, K., Habes, M., Klinger-König, J., Bülow, R., Völzke, H., & Grabe, H.J. (2020). A biomarker for Alzheimer's disease based on patterns of regional brain atrophy. *Frontiers in Psychiatry, 10*, 1-9. https://doi.org/10.3389/fpsyt.2019.00953

He, Z., Govindarajan, V., Yang, J., Abdel-Salam, M., Shaikh, Z.A., Por, L.Y., Alizadehsani, R., Tan, R.S., & Acharya, U.R. (2025). Automated Alzheimer's disease detection using active learning model with reinforcement learning and scope loss function. *NPJ Mental Health Research, 4*, 1-18. https://doi.org/10.1038/s44184-025-00158-2

Islam, J., & Zhang, Y. (2018). A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data. *In Proceedings of the International Conference on Brain Informatics*, pp. 1-10.

Jumaili, M.L.F., & Sonuç, E. (2025). ML-driven Alzheimer's disease prediction: A deep ensemble

modeling approach. *SLAS Technology, 32*, 1-15. https://doi.org/10.1016/j.slast.2025.100298.

Khan, Vinukonda, E. R., & Jagadesh, B. N. (2025). An integrated deep learning model for early and multi-class diagnosis of Alzheimer's disease from MRI scans. *Scientific Reports, 15*, pp. 1-21. https://doi.org/10.1038/s41598-025-01845-y.

Mabokela, K. R. (2019). A multilingual ASR of Sepedi-English code-switched speech for automatic language identification. *IEEE, In 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC),* pp. 1-8.

Mmadumbu, A.C., Saeed, F., Ghaleb, F., & Qasem, S.N. (2025). Early detection of Alzheimer's disease using deep learning methods. *Alzheimer's & Dementia, 21*, 1-12. https://doi.org/10.1002/alz.70175.

Mosconi, L., Brys, M., Glodzik-Sobanska, L., De Santi, S., Rusinek, H., & de Leon, M.J. (2007). Early detection of Alzheimer's disease using neuroimaging. *Experimental Gerontology, 42*, 129–138. https://doi.org/10.1016/j.exger.2006.05.016.

Mousavi, S.M., Moulaei, K., & Ahmadian, L. (2025). Classifying and diagnosing Alzheimer's disease with deep learning using 6735 brain MRI images. *Scientific Reports, 15*, 1-12. https://doi.org/10.1038/s41598-025-08092-1.

Muksimova, S., Umirzakova, S., Iskhakova, N., Khaitov, A., & Cho, Y.I. (2025). Advanced convolutional neural network with attention mechanism for Alzheimer's disease classification using MRI. *Computers in Biology and Medicine, 190*, 1-13. https://doi.org/10.1109/iSSSC50941.2020.9358837

Nestor, P. J., Scheltens, P., & Hodges, J. R. (2004). Advances in the early detection of Alzheimer's disease. *Nat. Med., 10*(Suppl), S34–S41. https://doi.org/10.1038/nm1433.

Noh, J.H., Kim, J.H., & Yang, H.D. (2023). Classification of Alzheimer's progression using fMRI data. *Sensors, 23*. https://doi.org/10.1007/978-3-030-05318-5_1.

Oraby, S., Emran, A., El-Saghir, B., & Mohsen, S. (2025). Hybrid of DSR-GAN and CNN for Alzheimer disease detection based on MRI images. *Scientific Reports*, *15*, 1-16. https://doi.org/10.18653/v1/w18-6116.

Roy, P., Ghosh, D., Sanyal, R., Madhu, N.R., & Dey, A. (2024). The Controversy Surrounding Drugs Against Neurodegenerative Disorders: Benefit or Harm? In: Pathak, S., Banerjee, A. (eds) Neuroprotective Effects of Phytochemicals in Brain Ageing. Springer, Singapore. pp. 373-386. https://doi.org/10.1007/978-981-99-7269-2_17

Saratxaga, C.L., Moya, I., Picón, A., Acosta, M., Moreno-Fernandez-de-Leceta, A., Garrote, E., & Bereciartua-Perez, A. (2021). MRI deep learning-based solution for Alzheimer's disease prediction. *J. Pers. Med., 11*, 1-22. https://doi.org/10.3390/jpm11090902

Shamrat, F.M.J.M., Akter, S., Azam, S., Karim, A., Ghosh, P., Tasnim, Z., Hasib, K.M., De Boer, F., & Ahmed, K. (2023). AlzheimerNet: An effective deep learning-based proposition for Alzheimer's disease stages classification from functional brain changes in magnetic resonance images. *IEEE Access, 11*, 16376-16386. https://doi.org/10.1109/ACCESS.2023.3244952.

Suchitra, S., Krishnasamy, L., & Poovaraghan, R. J. (2025). A deep learning-based early Alzheimer's disease detection using magnetic resonance images. *Multimed. Tools Appl., 84,* 16561–16582. https://doi.org/10.1109/ICACCI.2018.8554501

Sun, X., Zhu, J., Li, R., Peng, Y., & Gong, L. (2025). The global research of magnetic resonance imaging in Alzheimer's disease: a bibliometric analysis from 2004 to 2023. *Front. Neurol., 15*, 1-14. https://doi.org/10.3389/fneur.2024.1510522.

Zhou, J., Wei, Y., Li, X., Zhou, W., Tao, R., Hua, Y., & Liu, H. (2025). A deep learning model for early diagnosis of Alzheimer's disease combined with 3D CNN and Video Swin Transformer. *Scientific Reports,15*, 1-17. https://doi.org/10.1145/2600428.2609622.

How to cite this Article:

Shagufta Shakeel and Satya Prakash Yadav (2026). A Novel Hybrid Model for Multi-Class AD Classification using OASIS MRI Dataset. *International Journal of Experimental Research and Review, 50*, 25-40

DOI : https://doi.org/10.52756/ijerr.2026.v50.002

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.